

Certificate for ETD-repositories

Peter Schirmbacher

Director Computer and Mediaservice, Humboldt-University Berlin

Keywords: Certificate, Institutional Repositories, Requirements to repositories, NDLTD

ABSTRACT

During the last years a lot of universities worldwide created ETD-repositories. The community is increasing every day. In Germany all universities have their own ETD repository. In most cases the first step is already done. The next step should be a so-called quality control. Therefore we developed a special certificate for document and publication repositories by the German Initiative for Networked Information (DINI). I spoke about these activities last year during the ETD-conference in Kentucky. It is my suggestion to use these criteria also for repositories inside NDLTD. During the conference I intend to speak about the seven criteria for repositories and how we can introduce this certificate to the ETD community.

The seven criteria are:

- A policy for the repository
- Author support
- Legal aspects
- Authenticity and integrity
- Indexing
- Statistics
- Long term availability

1. INTRODUCTION

In the beginning of 2001 the German Initiative of Networked Information (DINI) established a special working group in the field of electronic publishing. This team consisting of 10 volunteers from eight different German universities has created recommendations for "Electronic publishing in higher education". These recommendations were the result of experiences in eight publishing or digital library projects, in most cases supported by the German Science Foundation (called DFG). With this paper, published in March 2002, we tried to exert influence on the development in Germany as a whole. (<http://www.dini.de/documents/DINI-EPUB-2002-03-10-E-V2.pdf>)

We discussed the problem if the new age of publishing requires a new culture of electronic publishing because many things and procedures are changing and demand some new basic rules and principles.

General rules or principles we recommended are the following:

1. Compliance with national and international rules and standards
 - It is important that every university repository tries to compliance the Open Archives Initiative.
 - To use Dublin Core Metadata Element Set is natural.
2. Avoidance of proprietary file formats
 - Using of SGML and XML should be the aim of long term activities
At Humboldt University we adopted conversion tools, adopted document type definitions and developed a special teaching and learning guide for electronic theses and dissertations.
3. Protection of authenticity and integrity

- For four years at some German universities every document, which was saved, get a time stamp and a digital signature in order to guarantee the authenticity and integrity.
4. Archiving and long term availability
- We gave some recommendations to hardware, software and orgware rules.

At the beginning of 2003 we started a new activity and sent a questioner to all German universities to ask them for more detailed information about their digital repositories' quality. We wanted to find out where the main mistakes are and difficulties in the process of building up repositories.

When we discussed the results we came to the conclusion that the best support for libraries could be a clear description of a document and publication repository. Everybody should know the necessary components for a repository to fulfill a minimum of standards and rules. What are the most important standards?

2. REQUIREMENTS TO REPOSITORIES

In the course of discussion we came to seven points, which are important for an institutional repository. In order to characterize a digital repository we are of the opinion that we need statements to

- Policy for the repository
- Author support
- Legal aspects
- Authenticity and integrity
- Indexing
- Statistics
- Long term availability

The criteria are split into two sections. In the first section so called ***minimum standards and requirements*** are specified, that must be met by the document and publication repositories or their operators. As DINI plans a continuous adaptation of the certificate to retain its validity in an ever changing environment, the second section lists ***recommendations***, which, from today's point of view, are likely to be future requirements for the certificate.

Policy

The operator of the document or publication repository formulates a policy regarding content and operation of the repository and makes this policy publicly available.

This policy must state rights and obligations of operators of a repository and of authors and/or editors.

The policy must contain:

- Provision of standards for the publications in the repository with regard to content and functional and technical quality
- A guarantee to archive for defined periods depending on content and functional and technical quality of the publication
- Procedures for the operation of the repository; definition of services that the operator of the document repository offers to authors and editors

Author support

It is recommended to initiate course programs for electronic publishing at the universities. To raise the level of attraction to and acceptance of electronic publishing, strategic advantages such as support services at all. must be developed and offered to authors and editors. DINI encourages the use of Document Type Definitions (DTD) according to international standards for publications, or the support of templates and the enforcement of style sheets for word processors and/or TeX/LaTeX.

Legal Aspects

This mentioned policy must be accepted by authors and editors as well as the university. This policy must state rights and obligations of operators of a repository and of authors and/or editors. The operator of the repository must enter contractual relations with authors according to the policy.

Minimum standards

- The operator of the repository must be permitted to publish the uploaded document onto the repository (right to store electronically, especially in databases, and make available to the public for individual retrieval, rendition on a monitor and print (online use) in part and/or as a whole)
- The operator of the repository must be permitted to forward the document to an archiving institution
- The operator of the repository must be permitted to alter the documents technically to secure long-term availability
- Exemption from liability must be formulated in a disclaimer

Recommendation

- List rights and privileges in metadata

Authenticity and Data Integrity

These criteria we have to divide in criteria for repositories and criteria for documents.

Minimum standards for servers

- Documentation of the technical system (It is expected of the repository to be available on a 24/7 basis)
- Back-up system to secure repository, metadata and documents
- Technically controlled and verifiable acceptance of documents

Recommendation for servers

- SSL certification

Minimum standards for documents

- Persistent identifier must be assigned
- A document with altered content must be treated as a new document (new persistent identifier)
- Archiving of authors' uploaded files in their original format

Recommendations for documents

- Application or procedure to account for the integrity of documents (e. g. a hash number)
- Advanced digital signature (§ 2 Abs. 2 SigG 2001¹)
- Rendering of archiving file formats with reference to the export of documents into long-term archiving facilities or institutions

Indexing

According Indexing, we have to divide the levels, subject indexing, Export of Metadata and Interfaces.

Subject indexing

Minimum standards

- Availability of a defined policy for indexing
- Verbal indexing with keywords or classificatory indexing

Recommendations

- Dewey Decimal Classification (DDC) according to usage in the German National Bibliography as the general classificatory indexing system for documents (v. DINI-OAI recommendations)
- At least one additional standardized system of verbal or classificatory indexing (general or subject specific, e. g. Schlagwortnormdatei, LoC, CCS, MSC, PACS etc.)

⁰ SigG –Signaturgesetz – German Signature Law

- Keywords in English
- Abstracts in German and English

Indexing - Export of Metadata

Minimum standards

- Metadata are available for free
- Unqualified Dublin Core (ANSI/NISO Z39.85–2001²)

Recommendations

- Qualified Dublin Core
- ONIX³
- Technical and/or archival metadata; print-on-demand data (e. g. METS⁴)
- Export to long-term archiving institutions or repositories
- Export into bibliographical databases (e. g. bibliographical management system, networks, OPACs)

Indexing - Interfaces

Minimum standards

- User interface to the web repository
- OAI PMH 2.0⁵

Recommendations

- Expansion of the OAI protocol to allow for exchange of complex metadata schemata
- Web-service interface (e. g. SOAP⁶)
- Z 39.50⁷

Logs and Statistics

Document-related access statistics, integrated into the document and publication repository, contain indications of the relevance of a publication in relation to other publications.

Access statistics of a document and publication repository are the qualitative, quantitative and technological basis for the evaluation of the repository.

Chronological progression within document-related access statistics can contain indications about the “up-to-dateness” of a publication throughout its life cycle.

Minimum standards

- Every individual repository must (within legal boundaries) log statistical data on access to the repository and to individual publications

Recommendation

- The log containing data on access to a publication should be added to the publication as dynamic metadata

Long-term Availability

During a period of five years following the publication date of a document, it is the operators' responsibility to ensure access to publications. This period must be defined in the document and publication repository's policy.

At the same time, the long-term availability of the electronic resources has to be ensured. In Germany it has to be organized in cooperation with the German National Library.

Minimum standards

- Persistent linking of metadata and documents (e. g. via a persistent identifier or storage of metadata and document in one single container)

¹ http://www.niso.org/standards/standard_detail.cfm?std_id=725

² <http://www.editeur.org/onix.html>

³ <http://www.loc.gov/standards/mets/>

⁴ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁵ <http://www.w3.org/TR/SOAP/>

⁶ http://www.niso.org/standards/resources/Z3950_Resources.html

- Definition within the policy (v. **Error! Reference source not found.**) of a minimum time span of availability of a document of no less than five years)

Recommendation

- Securing long-term availability where necessary through cooperation with a archiving institution

3. Quality control or certificate of repositories

By issuing this certificate, for the first time DINI introduces a quality control for document and publication repositories in Germany. A set of minimum requirements of a repository and its operator mandatory for modern scholarly communication is formulated. At the same time DINI makes recommendations highlighting predictable developments that might turn into future requirements. “DINI-Certificate for Document and Publication Repository” takes a step further by clearly describing criteria, which will guarantee repositories to be set up and operated according to national standards and international developments. Repositories fulfilling these criteria may be awarded a certificate, testifying to their quality.

The DINI office or an authorized working group is responsible for awarding the DINI Certificate – Document and Publication Repository in Germany. The seal of the certificate carries the year when it is awarded. The certification document acknowledges that the certified repository meets the minimum standards of a DINI-certified Document and Publication Repository. The operator of a document or publication repository applies at DINI for certification using a web form on the DINI website. This form contains the minimum standards as listed in the criteria. The applicant fills out the form therewith stating to what extent the criteria for certification are met by the repository to be certified. Further information about the repository may be added in designated fields in the form.

An authorized DINI working-group verifies the submitted data. The certification process should generally be completed within two months.

There is no time limit on the certificate. As the certificate shows the year it was awarded, reverification of criteria can be done at any time.

Currently we have about 12 institutional repositories of German universities which got a certificate.

In my opinion I believe a comparable certificate or at least the criteria for repositories should be an additional point of the ETD guide. It could help and give an orientation for those universities which start or reconstruct their repositories.